IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Applicant(s):	Santosh S. Rao, Gopal Sharma, Poonam Dhavale		
Assignee:	VERITAS Operating Corporation		
Title:	System And Method For Resolving Cluster Partitions In Out-Of-Band Storage Virtualization Environments		
Serial No.:	10/627,385	Filing Date:	July 25, 2003
Examiner:	Unassigned	Group Art Unit:	2184
Docket No.:	VRT0089US		

Austin, Texas
April 12, 2004

Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

PETITION TO MAKE SPECIAL UNDER 37 CFR §1.102(d)

Dear Sir:

The applicants hereby petition pursuant to 37 CFR §1.102(d) and MPEP § 708.02(VIII) to make the above-identified application special. Please charge Deposit Account No. 502306 for the fee of $130.00 for this petition as set forth in 37 CFR §1.17(h).

Should the Office determine that all the claims presented are not obviously directed to a single invention, the applicants will make an election without traverse as a prerequisite to the grant of special status.

The applicants respectfully submit that a pre-examination search has been performed by a professional search firm in the following classes/subclasses:

Class	Subclasses
G06F	11/00, 11/14, 11/18, 12/00, 17/30
709	205, 220, 232
714	4, 6, 13

Enclosed are copies of the following references which are presently believed to be, from among those made of record in the accompanying Information Disclosure

Statement and any previously filed Information Disclosure Statement, the most closely related to the subject matter encompassed by the claims:

US 6,658,587	US 6,002,851	WO 98/33121
WO 03/027903	EP 1274012	EP 0810526

Detailed Discussion of the References

U.S. Patent 6,002,851 (Basavaiah) and related PCT Application WO 98/33121 (Jardine) both disclose techniques for handling split-brain conditions in multi-processor systems having a plurality of processors.

For example, Jardine discloses a protocol to determine the group of processors that will survive communications faults in a multiprocessor system. Processors embodying Jardine's invention construct a connectivity matrix on the initiation of a regroup operation, typically triggered by some processor failure or inter-processor communication failure. The connectivity information is used to ensure that all the processors in the final group that survives can communicate with all other processors in the group. One or more processors may halt to achieve this characteristic. See, e.g., page 17, lines 23-30.

Referring to Figure 3 of Jardine, a two processor multi-processor system 300 is shown. System 300 is said to be fully connected because the two processors 1, 2 are in communication with each other. When a fault occurs that divides a system such as system 200 (Figure 2) into a graph such as graph 400 (Figure 4), the group of processors 1, 3, 4 and 5 is fully connected, and the group of processors 1, 2 and 5 is fully connected. The processors of graph 400 all enter a regroup operation on the detection of the communication failures. According to the teachings of Jardine, in order to avoid split-brain problems and to maintain a fully connected multiprocessor system, processor 2 halts operations, while each of the processors 1, 3, 4 and 5 continues operations. See, e.g., page 22, line 19 to page 23, line 7.

To accomplish the regrouping or processors, Jardine teaches the selection of a tie-breaker processor. Referring to Figure 1, Jardine teaches that one of the processors 112 has a special role in the regroup process, and is designated the tie breaker. The split-brain avoidance process favors the tie breaker processor in case of ties. Further, the node pruning process used to ensure full connectivity between all surviving processors is run on the tie-breaker processor. This process also favors the tie breaker in case of large numbers of connectivity failures. Each of the processors 112 of the multi-processor system uses network 114 for broadcasting "IamAlive" messages at periodic intervals.

Approximately every 2.4 seconds, each processor 112 checks to see what IamAlive messages it has received from its companion processors. When a processor fails to receive an IamAlive message from another processor (e.g., 112b) that it knows to have been a part of the system at the last check, the checking processor initiates a regroup operation by broadcasting a "Regroup" message. In effect, a regroup operation is a set of chances for the processor 112b from which an IamAlive message was not received to convince the other processors 112 that it is in fact healthy. Processor 112b's failure to properly participate in the regroup operation results in the remaining processors 112 ignoring any further message traffic from processor 112b, should it send any. The other processors exclude processor 112b from the system. See, e.g., page 24, line 25 to page 25, line 26.

Thus, in order to resolve a split-brain condition among processors in a multi-processor system, Jardine and Basavaiah teach the creation of a connectivity matrix for use in determining which processors remain connected each other, and deadlock resolution performed by a selected one of the surviving processors. In contrast, the applicants' claims generally address the split-brain problems in computer system clusters, not within multiprocessor systems. More specifically, Jardine and Basavaiah neither teach nor suggest: (1) providing a coordinator *virtual* device for use in split-brain resolution; (2) attempting to gain control of the coordinator virtual device; or (3) removing a node from the computer system cluster when such attempting is unsuccessful, all as required by claim 1, and generally required by claims 15 and 28. Accordingly, the applicants respectfully submit that claims 1-31 are allowable over Jardine and Basavaiah.

U.S. Patent No. 6,658,587 (Pramanick) discusses the use of quorum algorithms and quorum devices to resolve split-brain problems in computer system clusters. Referring to Figure 1, if a communications failure occurs between node 106 and the other nodes in the cluster and each node has one vote that can be used to resolve split-brain conditions, then a simple quorum algorithm can be used. Since nodes 102, 104 and 108 are operating properly and are in communication with one another, a simple quorum algorithm would count one vote for each of these devices, against one vote for node 106.

Since 3 > 1, the subcluster comprising nodes 102, 104 and 108 attains majority vote count and this simplified quorum algorithm excludes node 106 from accessing shared disk 110. However, if the split-brain event leads to the same number of nodes in each subcluster, then no subcluster attains majority vote count and this relatively simple quorum algorithm fails. A quorum device, is a hardware device shared by two or more nodes within the cluster that contributes votes used to establish a quorum and avoid this problem. Pramanick further notes that quorum devices are commonly shared disks, and most majority vote count quorum algorithms assign the quorum device a number of votes which is one less than the number of connected quorum device ports. See, e.g., column 3, lines 1-51.

Pramanick further teaches that certain SCSI-3 commands (Persistent Group Reservation features, or PGRs) allow a host node to make a disk reservation that is persistent across power failures and bus resets, and permit group reservation allowing all nodes in a running cluster to have concurrent access to the disk while disallowing access to nodes not in the cluster. To that end, Pramanick teaches that quorum disks featuring PGRs are useful. Because not all devices, or even all SCSI devices support the PGRs, Pramanick teaches a technique for emulating the PGRs on non-PGR compliant devices. See, e.g., column 4, lines 3-17, and column 4, line 60 through column 5, line 18.

Thus, while Pramanick teaches the use of quorum devices to resolve disputes among nodes surviving a split-brain condition, Pramanick neither teaches nor suggests: (1) providing a coordinator virtual device for use in split-brain resolution; (2) attempting to gain control of the coordinator virtual device; or (3) removing a node from the computer system cluster when such attempting is unsuccessful, all as required by claim 1, and generally required by claims 15 and 28. Accordingly, the applicants respectfully submit that claims 1-31 are allowable over Pramanick. European Applications 1274012 (Shirriff) and 0810526 (Satyanarayanan) similarly describe quorum device techniques and vote gathering (See, e.g., respective abstracts), but do not teach or suggest the applicants' claim limitations.

PCT Application 03/027903 (Callahan) describes a system and method for providing a multi-node environment where each node has an operating system independent from the operating system of each other node, and where the nodes share a storage and an interconnect allowing the nodes to directly access the storage. See, e.g., abstract.

More specifically, Callahan describes accommodating situations where membership in the system has changed. Referring to Figures 7A and 7B, Callahan teaches determining whether the membership of a cluster has changed (702) due to, for example, system failure. If cluster membership has changed, then new locks are no longer granted (710) and it is then determined whether there is an administrator (ADM) in this cluster (712). If there is no administrator in this cluster, then one of the members of the cluster is elected as administrator (714). The administrator verifies that the other servers in the cluster are part of this storage area membership (720). Step 720 accommodates both when all of the servers are part of the cluster, or when there are servers outside the cluster. If nodes are outside the cluster then the administrator excludes (fences) those servers to prevent corruption of data on shared storage. Servers that successfully gain membership to the network cluster are then allowed access to the shared storage and are then part of the SAN membership. All cluster non-members are then excluded and all cluster members are allowed into the shared storage group (722). See, e.g., page 15, line 17 through page 17, line 20.

Thus, although Callahan discloses a system that uses some mechanism to separate cluster members from non-members, Callahan neither teaches nor suggests any specific techniques to accomplish this. Moreover, Callahan neither teaches nor suggests: (1) providing a coordinator virtual device for use in split-brain resolution; (2) attempting to gain control of the coordinator virtual device; or (3) removing a node from the computer system cluster when such attempting is unsuccessful, all as required by claim 1, and generally required by claims 15 and 28. Accordingly, the applicants respectfully submit that claims 1-31 are allowable over Callahan.

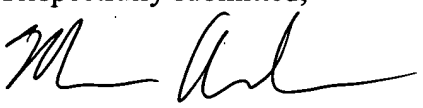
Conclusion

In summary, the applicants respectfully submit that none of the references located during the pre-examination search, or otherwise made of record by the applicants, teaches or suggests (at least): (1) providing a coordinator virtual device for use in split-brain resolution; (2) attempting to gain control of the coordinator virtual device; or (3) removing a node from the computer system cluster when such attempting is unsuccessful, all as required by claim 1, and generally required by claims 15 and 28.

Accordingly, the applicants respectfully request that this petition be granted, and that the present application receive expedited examination. Should any issues remain that might be subject to resolution through a telephonic interview, the Office is requested to telephone the undersigned.

Express Mail Label No:
EV 304739152 US

Respectfully submitted,



Marc R. Ascolese
Attorney for Applicant(s)
Reg. No. 42,268
512-439-5085
512-439-5099 (fax)